P25268

Video Mixing Apparatus and Method of Mixing Video

Field of the Invention

The present invention relates to a video mixing apparatus carrying out a chromakey process and a method of the chromakey process.

Background of the Invention

Related arts are disclosed by USP4,533,937 and JP Patented No. 2,556,810. First, the prior art disclosed by USP4,533,937 is described hereinafter. Fig. 12 is a block diagram illustrating a first example of a video mixing apparatus carrying out a conventional chromakey process.

In this first example, a component of a foreground object is taken out from a source video signal, and mixed with background video signal. The source video signal is generated by shooting the foreground object in front of a monochromatic screen. This is known as a chromakey process.

As shown in Fig. 12, this video mixing apparatus performs as follows: Key signal generator 1101 generates key signal K using source video signal Vs and a screen reference color. Screen reference color memory 1102 stores screen reference color Vr designated by a user. Mixing processor 1103 mixes source video signal Vs with background video signal Vz based on key signal K and screen reference color Vr, thereby outputting mixed video signal Vm.

Reference marks Vs, Vz, Vr and Vm are three-dimensional color vectors of which respective luminance component Y, blue color difference component Cb, and red color difference component Cr are (Sy, Scb, Scr), (Zy, Zcb, Zcr), (Ry, Rcb, Rcr), (My, Mcb, Mcr), and reference mark K is a scalar value taking $0 \leqq K \leqq 1$.

Key signal generator 1101 generates key-signal K, for instance, as follows:

Formula (1) indicates distance "d" ($d \geqq 0$) on Cb – Cr plane between source video signal Vs and screen reference color Vr. Output signal K is a function of input signal "d" and is saturated by base-clip-level "db" and peak-clip-level "dp". This is shown in formula (2).

$$d = \sqrt{(Scb - Rcb)^2 + (Scr - Rcr)^2} \qquad (1)$$

$$K = \mathrm{Clip}(d, db, dp) \qquad (2)$$

where function "Clip" in formula (2) produces an output saturated with respect to the input signal at base clip level "b" and peak clip level "p" as shown in Fig. 13. Function "Clip" carries out the following calculations:

i) When $d \leq b$, $\mathrm{Clip}(d,b,p) = 0$ (3a)

ii) When $b < d < p$, $\mathrm{Clip}(d,b,p) = (d - b)/(p - b)$ (3b)

iii) When $p \leq d$, $\mathrm{Clip}(d,b,p) = 1$ (3c)

An example case, where base-clip-level "db" and peak-clip-level "dp" take constant values, is described hereinafter. Key signal generator 1101, as shown in Fig. 14, generates key signal K based on a distribution where two circles having respective radiuses "db" and "dp" draw boundaries. In other words, regarding distance "d", the following three cases are available:

i) When "d" is inside 140 of the circle having radius "db", $K = 0$.

ii) When "d" is outside 142 of the circle having radius "dp", $K = 1$.

iii) When "d" is at a place other than the above two cases,

$0 < K = (d - db) / (dp - db) < 1$.

As such, in this first example, key signal generator 1101 determines key signal K using only a color difference value of the source video signal.

Fig. 15 shows an example of mixing processor 1103, which mixes source video signal Vs with background video signal Vz based on key signal K, thereby outputting mixed video signal Vm. Adder 1401 finds a component of foreground object by subtracting (1 – K)Vr from source video signal Vs, where (1

– K)Vr is a screen component calculated by multiplier 1402. Another adder 1404 outputs mixed video signal Vm by adding (1 – K)Vz to foreground object component Vf, where (1 – K)Vz is a background video component calculated by another multiplier 1403.

In the structure shown in Fig. 14, mixing processor 1103 outputs mixed video signal Vm according to the following formulas (4):

$$\text{Y component: } My = Sy - (1 - K)Ry + (1 - K)Zy \tag{4a}$$

$$\text{Cb component: } Mcb = Scb - (1 - K)Rcb + (1 - K)Zcb \tag{4b}$$

$$\text{Cr component: } Mcr = Scr - (1 - K)Rcr + (1 - K)Zcr \tag{4c}$$

Next, the prior art disclosed by JP Patented No. 2,556,810 is described hereinafter. Fig. 16 is a block diagram illustrating a second example of the conventional video mixing apparatus. This apparatus also carries out the chromakey process. This second example differs from the first one in the following point: Key signal generator 1501 outputs two key signals Kc (color-canceling key signal) and Km (mixing key signal). Mixing processor 1503 calculates the mixing based on the two key signals Kc, Km. In this example, the mixing is carried out according to the following formulas:

$$\text{Y component: } My = Sy - (1 - Km)Ry + (1 - Km)Zy \tag{5a}$$

$$\text{Cb component: } Mcb = Scb - (1 - Kc)Rcb + (1 - Km)Zcb \tag{5b}$$

$$\text{Cr component: } Mcr = Scr - (1 - Kc)Rcr + (1 - Km)Zcr \tag{5c}$$

As formula (5) tells, in this example, color-canceling key signal Kc used for subtracting a color difference component can have a value different from mixing process key signal Km. Thus, it can be set that adding of a background video signal component is carried out in a narrow region on a color space and weakening of a screen color is carried out in a rather greater region. Therefore, at a portion where the screen color mixes into the foreground object, the screen

color can be weakened without any transparency in the foreground object.

However, the video mixing apparatus of this example generates a key signal using only color difference information of the source video signal, thus, it sometimes cannot separate properly the foreground object component from the screen component.

Hereinafter described is a way of generating a key signal from the source video shown in Fig. 17. Respective pixels constituting the source video shown in Fig. 17 are projected to a plane vertical to a color difference plane. Fig. 18 shows this projection. The pixels constituting the foreground object are distributed in the vicinity of region F 1701 in Fig. 18, while the pixels constituting the screen are distributed in the vicinity of region X 1702. Between these two regions, the pixels constituting the edge of foreground object are distributed, and the pixels are mixed with both the components of foreground object and the screen.

In this example, it is difficult to set a proper edge (boundary) for generating a key signal. For instance, when a boundary face—separating a region of $K = 1$ and another region of $K < 1$—is set as a boundary face "a" denoted with 1801 in Fig. 19, the component of foreground object is weakened. On the other hand, when the boundary face is set as boundary face "b" denoted with 1802 in Fig. 19, the screen color invades the edge in the mixed video.

In the way discussed above, when the mixing process is carried out by the calculations shown in formulas (4), noises (uneven lighting, scratches, stains, and the like) in the screen appear in the mixed video, thereby lowering the picture quality. This is because vector Vr having a constant value is used as a subtracting vector for removing the screen component.

Basically, a pixel—receiving key signal $K = 0$ at the key signal generator—is to be mixed video signal Vm = background video signal Vz. In other words,

regarding a luminance component, the term of $Sy - (1 - K)Ry$ in (4a) should take "0" (zero) and $My = Zy$ should be satisfied. However, since the screen has some noises, $Sy \neq Ry$, thus error of $Sy - Ry$ appears in the mixed video.

Summary of the Invention

The present invention addresses the problems discussed above, and aims to provide a video mixing apparatus which can separate a foreground object from a screen, and a method of the same. According to the present invention, the foreground object and the screen are properly separated taking a luminance component into consideration for reserving more discretion at separation. Further, according to the present invention, a key signal is generated based on a boundary face set in three-dimensional space.

The present invention aims to provide a video mixing apparatus which can remove a screen component including noises.

The video mixing apparatus of the present invention comprises a key signal generator and a screen signal generator. The key signal generator generates a key signal based on a position of a screen reference color in a key signal distribution formed by two ovals set in the three-dimensional space.

The key signal generator sets an oval body in key signal distribution in the three-dimensional space including the luminance component. Further, the key signal generator generates a key signal based on a position of a screen reference color in the key signal distribution. Thus, the key signal generator can separate properly the foreground object from the screen.

A screen signal generator outputs a screen signal to respective pixels of the source video signal, so that a screen component including noises is removed.

Brief Description of the Drawings

Fig. 1 is a block diagram illustrating a structure of a video mixing apparatus in accordance with a first exemplary embodiment of the present invention.

Fig. 2 shows relation of input and output of function Clip (d, b, p).

Figs. 3A and 3B show distribution of key signals in color space, the key signals are generated by a key signal generator in accordance with the first exemplary embodiment of the present invention.

Fig. 4 shows an example of a source video signal.

Fig. 5 shows respective pixels projected on a plane vertical to a color difference component plane, the pixels constitute the source video signal.

Fig. 6 shows a boundary face specifying distribution of key signals generated by a key signal generator and a cross section of pixels constituting a screen.

Fig. 7 shows another distribution of the key signal generated by the key signal generator in accordance with the first exemplary embodiment of the present invention.

Fig. 8 is a block diagram illustrating a structure of a video mixing apparatus in accordance with a second exemplary embodiment of the present invention.

Figs. 9A and 9B illustrate screen signals generated by a screen signal generator in accordance with the second exemplary embodiment of the present invention.

Fig. 10 is a block diagram illustrating a structure of a video mixing apparatus in accordance with a third exemplary embodiment of the present invention.

Fig. 11 shows distribution of key signals generated by the key signal

generator in accordance with the third exemplary embodiment of the present invention.

Fig. 12 is a block diagram illustrating a structure of a conventional video mixing apparatus.

Fig. 13 shows relation of input and output of function Clip (d, b, p).

Fig. 14 shows distribution of key signals generated by the conventional key signal generator.

Fig. 15 shows a circuit structure generating mixed video signal Vm using source video signal Vs and background video signal Vz.

Fig. 16 is a block diagram illustrating another structure of a conventional video mixing apparatus.

Fig. 17 is a source video signal.

Fig. 18 shows respective pixels constituting the source video signal, the pixels are projected on a plane vertical to a color difference component plane.

Fig. 19 is a cross section of respective pixels constituting a screen, a foreground object and a boundary face, these three specify a distribution of a key signal generated by the conventional key signal generator.

Description of the Preferred Embodiments

Exemplary embodiments of the present invention are demonstrated hereinafter with reference to the accompanying drawings.

Exemplary Embodiment 1

Fig. 1 is a block diagram illustrating a structure of a video mixing apparatus in accordance with a first exemplary embodiment of the present invention. In Fig. 1, key signal generator 101 generates key signal K using source video signal Vs and screen reference color Vr.

Screen reference color memory 102 stores screen reference color Vr

designated by a user.

Mixing processor 103 mixes source video signal Vs with background video signal Vz based on key signal K and screen reference color Vr, thereby outputting mixed video signal Vm.

Reference marks Vs, Vz, Vr and Vm are three-dimensional color vectors of which respective luminance component Y, blue color difference component Cb, and red color difference component Cr are (Sy, Scb, Scr), (Zy, Zcb, Zcr), (Ry, Rcb, Rcr), (My, Mcb, Mcr), and reference mark K is a scalar value taking $0 \leqq K \leqq 1$.

Mixing processor 103 outputs mixed video signal Vm following the calculations shown by formulas (4) described in the first example of the conventional method. Key signal generator 101 outputs key signal K using the received source signal Vs and screen reference color Vr and following the calculations shown by formulas (6) and (7) below:

$$d1 = \sqrt{t^2 (Sy - Ry)^2 + (Scb - Rcb)^2 + (Scr - Rcr)^2} \qquad (6)$$

$$K = \mathrm{Clip}(d1, Ar, Ar + Aw) \qquad (7)$$

Function "Clip" shown by formula (7) carries out the calculation shown by formula (3). As shown in Fig. 2, a value saturated at base-clip-level "b" and peak-clip-level "p" is output responding to input signal "d". Parameters "Ar", "Aw" and "t" are set by a user, and Ar > 0, Aw > 0, t > 0.

A distribution of key signal K calculated by formulas (6) and (7) is described hereinafter with reference to Fig. 3. Fig. 3A shows the distribution of key signal K in a three-dimensional space having coordinates axes of Y, Cb, Cr.

In Fig. 3, Vs indicates a position of the source video signal supplied in the three-dimensional space, and Vr indicates a position of the screen reference color designated by the user in the three-dimensional space. First oval body E0 has dimensions of center = Vr, a length of shorter axis = Ar, a length of longer axis = Ar/t. Second oval body E1 has dimensions of center = Vr, a length of shorter

axis = Ar + Aw, a length of longer axis = (Ar + Aw)/t.

The following formulas (8), (9) represent the first and the second oval body.

$$t^2 (Y - Ry)^2 + (Cb - Rcb)^2 + (Cr - Rcr)^2 = Ar^2 \qquad (8)$$

$$t^2 (Y - Ry)^2 + (Cb - Rcb)^2 + (Cr - Rcr)^2 = (Ar + Aw)^2 \qquad (9)$$

where, screen reference color Vr = (Ry, Rcb, Rcr)

In Fig. 3A, source video signal Vs = (Sy, Scb, Scr). Fig. 3B is a cross section of the oval body taken along a plane passing through Vs, Vr and vertical to Cb – Cr plane.

As shown in Fig. 3B, "d1" calculated by formula (6) passes through Vs and represents a length of a shorter axis of an oval body which has a center Vr shown in Fig. 3B and satisfies the following formula:

shorter axis length／longer axis length = t ($t > 0$)

Formula (7) finds key signal K by providing "d1" with a saturation process while Ar is set at a base-clip-level and Ar + Aw is set at a peak-clip-level. In other words, key signal K is determined based on key signal distribution of which equivalent planes form oval bodies having center Vr (screen reference color). As a result, the following key signal distribution shown in Fig. 3A is formed:

i) When Vs is inside oval body E0, K = 0.

ii) When Vs is outside oval body E1, K = 1.

iii) When Vs is between oval bodies E0 and oval E1, $0 < K < 1$.

A boundary face can be set for separating properly the foreground object from the screen by using oval bodies E0 and E1 shown in Fig. 3A.

Here is one example showing the production of a key signal from a source video shown in Fig. 4. Fig. 5 shows respective pixels—constituting the source video shown in Fig. 4—projected on a plane vertical to a color difference component plane. In Fig. 5, the pixels constituting the foreground object are

distributed in the vicinity of region F 501, while the pixels constituting the screen are distributed in the vicinity of region X 502. Between these two regions, the pixels constituting the edge of foreground object are distributed, and the pixels are mixed with both the components of foreground object and the screen. In this case, if oval bodies E0 and E1 shown in Fig. 3A are used, boundary faces c 603 and d 604 can be set, so that region F 501 and X 502 are properly separated. Meanwhile, boundary face c 603 is an oval body surrounding region F 501, and boundary face d 604 is an oval body surrounding boundary face c 603.

A user can arbitrarily set parameters "t", Ar and Aw. These parameters specify a shape of the oval body. Therefore, a proper boundary face can be set responsive to characteristics of the source video signal even when the distribution shown in Fig. 5 is not formed.

As such, the video mixing apparatus in accordance with this embodiment can set a boundary with a plane of an oval body, so that a key signal—separating a component of foreground object from a screen component—can be generated. As a result, a mixed video without lowering video quality can be supplied.

In this embodiment, mixed calculations shown by formulas (4) are used, but not limited to formulas (4), and multiplication key type mixed calculations shown by formulas (10) can be used with the same result.

Y component : $My = KSy + (1 - K)Zy$ (10a)

Cb component : $Mcb = KScb + (1 - K)Zcb$ (10b)

Cr component : $Mcr = KScr + (1 - K)Zcr$ (10c)

In this embodiment, the key signal distribution is formed by two oval bodies which share a common center Vr (screen reference color) and also have the same a ratio of shorter axis length vs. longer axis length. It is not limited to these oval bodies, but the key signal distribution can be formed by another two

oval bodies having different centers and the different ratios of shorter axis length vs. longer axis length. In this case, vector VrVs 52 starting from screen reference color Vr toward source video signal Vs crosses with oval body E0 at point "b", and distance "e0" between point "b" and Vr is a base-clip-level. Vector VrVs 52 crosses with oval body E1 at point "a", and distance "e1" between point "a" and Vr is a peak-clip-level. Then calculations are carried out so that an output responsive to distance "d" between Vr and Vs is saturated at the base-clip-level and peak-clip-level with respect to distance "d", thereby generating the key signal K.

In this case, the key signal distribution is formed to satisfy the following formulas:

When "d" $\leqq$ "e0", then K = 0.

When "d" $\geqq$ "e1", then K = 1.

When "e0" < "d" < "e1", then 0 < K < 1.

This case also brings the same result as discussed previously.

In this embodiment, a video mixing apparatus, which generates a key signal and mixes video using the key signal, is described. The key signal generated can be supplied to a video editing device and the like for further utilization.

(Exemplary Embodiment 2)

In this embodiment, a video mixing apparatus—having a screen signal generator for generating a screen signal—is demonstrated.

Fig. 8 is a block diagram illustrating a structure of the video mixing apparatus in accordance with the second exemplary embodiment of the present invention. In Fig. 8, key signal generator 801 generates key signal K using source video signal Vs and screen reference color Vr. Screen reference color

memory 802 stores screen reference color Vr designated by a user. Screen signal generator 803 generates screen signal Vx using source video signal Vs and screen reference color Vr. Mixing processor 804 mixes source video signal Vs with background video signal Vz based on key signal K and screen signal Vx, thereby generating mixed video signal Vm.

Reference marks Vs, Vz, Vr, Vx and Vm are three-dimensional color vectors of which respective luminance component Y, blue color difference component Cb, and red color difference component Cr are (Sy, Scb, Scr), (Zy, Zcb, Zcr), (Ry, Rcb, Rcr), (Xy, Xcb, Xcr), (My, Mcb, Mcr), and reference mark K is a scalar value taking $0 \leqq K \leqq 1$.

Key signal generator 801 outputs key signal K following the calculations shown by formulas (6) and (7) described in the first embodiment. Screen signal generator 803 outputs screen signal Vx using the received source signal Vs and screen reference color Vr and following the calculations shown by formulas (11) and (12) below.

$$d2 = \sqrt{t^2 (Sy - Ry)^2 + (Scb - Rcb)^2 + (Scr - Rcr)^2} \qquad (11)$$

$$\text{i) When } d2 \leq Ar, \qquad Vx = Vs \qquad (12a)$$

$$\text{ii) When } d2 \geq Ar, \qquad Vx = Vr + (Ar/d2)(Vs - Vr) \qquad (12b)$$

Parameters “Ar”, “Aw” and “t” are set by a user, and Ar > 0, Aw > 0, T > 0.

In other words, as shown in Fig. 9A, when key signal K is output so that source video signal Vs is inside first oval body E0 and K = 0 is satisfied in the key signal generator, screen signal generator 803 outputs source video signal Vs as it is.

As shown in Fig. 9B, when a key signal is supplied so that source video signal Vs is outside oval E0 and K > 0 is satisfied in the key signal generator, screen signal generator 803 outputs a coordinates value of point “c” as screen signal Vx, where point “c” is a cross point of vector VrVs 62 starting from screen

reference color Vr toward source video signal Vs and oval body E0.

Mixing processor 804 carries out a mixing process, based on key signal K supplied from key signal generator 801 and screen signal Vx supplied from screen signal generator 803, following formulas (13).

$$\text{Y component} \quad : My = Sy - (1 - K)Xy + (1 - K)Zy \qquad (13a)$$

$$\text{Cb component} \quad : Mcb = Scb - (1 - K)Xcb + (1 - K)Zcb \qquad (13b)$$

$$\text{Cr component} \quad : Mcr = Scr - (1 - K)Xcr + (1 - K)Zcr \qquad (13c)$$

As such, according to this second embodiment, Vx = Vs is output to the pixels to which key signal generator 801 supplies K = 0. A luminance component, i.e., the term of $Sy - (1 - K)Xy$ in formula (13a), takes 0 (zero). Therefore, the screen component is completely removed.

In this embodiment, the key signal distribution is formed by two oval bodies which share a common center Vr and also have the same a ratio (t) of shorter axis length vs. longer axis length. It is not limited to these oval bodies, but the key signal distribution can be formed by another two oval bodies having different centers and different ratios of shorter axis length vs. longer axis length. In this case, Vx = Vs is supplied to source video signal Vs which receives key signal K (K = 0), while a coordinates value of the cross point of vector VrVs starting from Vr toward Vs and oval body E0 is supplied as screen signal Vx to source video signal Vs which receives key signal K (K > 0).

(Exemplary Embodiment 3)

A video mixing apparatus, of which key signal generator generates two key signals, i.e., a color-canceling key signal and a mixing key signal, is demonstrated in the third embodiment. This apparatus carries out mixing calculations based on the two key-signals and a screen reference color.

Fig. 10 is a block diagram illustrating a structure of the video mixing

apparatus in accordance with the third exemplary embodiment. In Fig. 10, key signal generator 1001 generates color-canceling key signal Kc and mixing key signal Km using source video signal Vs and screen reference color Vr. Screen reference color memory 1002 stores screen reference color Vr designated by a user. Mixing processor 1003 mixes source video signal Vs with background video signal Vz based on key signals Kc and Km.

Reference marks Vs, Vz, Vr and mixed video signal Vm are three-dimensional color vectors of which respective luminance component Y, blue color difference component Cb, and red color difference component Cr are (Sy, Scb, Scr), (Zy, Zcb, Zcr), (Ry, Rcb, Rcr), (My, Mcb, Mcr), and reference mark Kc and Km are scalar values taking $0 \leqq Kc, Km \leqq 1$.

Key signal generator 1001 outputs color-canceling key signal Kx and mixing key signal Km following formulas (14) – (16).

$$d3 = \sqrt{t^2 (Sy - Ry)^2 + (Scb - Rcb)^2 + (Scr - Rcr)^2} \qquad (14)$$

$$Kc = Clip(d3, Ar, Ar + Aw1) \qquad (15)$$

$$Km = Clip(d3, Ar, Ar + Aw1 + Aw2) \qquad (16)$$

where parameters "Ar", "Aw1", "Aw2" and "t" are set by a user, and $Ar > 0$, $Aw1 > 0$, $Aw2 > 0$, $t > 0$.

Distributions of color-canceling key signal Kc and mixing key signal Km, both the signals are calculated by formulas (14) – (16), are described with reference to Fig. 11.

In this embodiment, three oval bodies form the distribution of the key signals, i.e., the first oval E0, second oval E1 and third oval E2 share a common center Vr and the same ratio "t" of shorter axis length vs. longer axis length, and E1 surrounds E0, E2 surrounds E1. As shown in Fig. 11, oval bodies E0 and E1 having respective shorter axis lengths "Ar" and "Ar + Aw1" specify the distribution of mixing key signal Km. Oval body E0 and oval body E2 having a

shorter axis length "Ar + Aw1 + Aw2" specify the distribution of color-canceling key signal Kc. In other words, what the user should do is just to set parameter Aw2 which specifies oval body E2 besides parameters "Ar", "Aw", and "t" which specify the distribution of mixing key signal Km.

As such, the video mixing apparatus in accordance with this third embodiment specifies the key-signal-distribution with the oval body planes defined in the three-dimensional space. Therefore, a boundary, which can separate properly a foreground object component from a screen component, can be set, and as a result, videos can be mixed with less degradation of the picture quality. In addition to the parameters for two oval bodies E0 and E1, which specify the distribution of mixing key signal Km, a parameter for oval body E2 surrounding the two oval bodies is only to be set, thereby specifying the distribution of color-canceling key signal Kc. The user, therefore, can operate the apparatus with ease.

Mixing processor 1003 carries out a mixing process—based on key signal K supplied from key signal generator 1001 and screen signal Vx supplied from screen signal generator 1002—following formulas (5), thereby generating mixed video signal Vm.

In this embodiment, mixed calculations shown by formulas (5) are used, but not limited to formulas (5), and multiplication key type mixed calculations shown by formulas (17) can be used with the same result.

$$\text{Y component}: \quad My = KmSy + (1 - Km)Zy \qquad (17a)$$

$$\text{Cb component}: Mcb = Km(Scb - (1 - Kc)Rcb) + (1 - Km)Zcb \qquad (17b)$$

$$\text{Cr component}: Mcr = Km(Scr - (1 - Kc)Rcr) + (1 - Km)Zcr \qquad (17c)$$

In this embodiment, the key signal distribution is formed by three oval bodies which share a common center and also have the same ratio of shorter axis length vs. longer axis length. It is not limited to these oval bodies, but the key

signal distribution can be formed with the same result by another three oval bodies having different centers and different ratios of shorter axis length vs. longer axis length.

In this case, the first and second oval bodies specify the distribution of the mixing key signal, and the second and third oval bodies specify the distribution of the color-canceling key signal.

As discussed above, according to the present invention, the key signal distribution is formed based on the oval body's face defined in the three-dimensional space. The key signal—properly separating the foreground object component from the screen component—is selected with ease and generated. As a result, mixing videos with less degradation of picture quality is obtainable.

A screen signal suitable to each pixel of a source video signal is generated and used as a vector subtracting from the source signal. Therefore, a screen component including noises can be removed.